Energy Economics 137 (2024) 107768

Contents lists available at ScienceDirect

Energy Economics

journal homepage: www.elsevier.com/locate/eneeco

Returns to solar panels in the housing market: A meta learner approach

Elias Asproudis [a], Cigdem Gedikli [a,*], Oleksandr Talavera [b], Okan Yilmaz [a]

[a] Swansea University, United Kingdom
[b] University of Birmingham, United Kingdom

ARTICLE INFO

JEL codes:
R21
R31
Q42
Q5

Keywords:
Solar panels
Residential housing market
Sale prices
Machine-learning
Meta-learners

ABSTRACT

This paper aims to estimate the returns to solar panels in the UK residential housing market. Our analysis applies a causal machine learning approach to Zoopla property data containing about 5 million observations. Drawing on meta-learner algorithms, we provide strong evidence documenting that solar panels are directly capitalized into sale prices. Our results point to a selling price premium above 6% (range between 6.1% to 7.1% depending on the meta-learner) associated with solar panels. Considering that the average selling price is £230,536 in our sample, this corresponds to an additional £14,062 to £16,368 selling price premium for houses with solar panels. Our results are robust to traditional hedonic pricing models and matching techniques, with the lowest estimates at 3.5% using the latter. Despite the declining trend, the additional analyses demonstrate that the positive premium associated with solar panels persists over the years.

1. Introduction

Solar photovoltaic (PV) panels stand out as an important renewable energy alternative in addressing climate change and reducing greenhouse emissions. They offer households access to self-generated electricity and feed-in tariff revenue (FiT), the benefits that are particularly worthwhile in today's world facing soaring energy costs. The rising popularity of solar PV panels over the years is not surprising and reflects the increased demand for less costly options and public awareness in valuing environmentally friendly choices, alongside the government incentives promoting alternative green energy technologies.[1] The multiple benefits solar PV panels offer are likely to be capitalized into how solar panels are valued in the housing market. In this paper, we seek to explore this by looking into whether properties with solar panels are associated with higher sale prices than those without solar panels.

The relationship between various housing attributes and rental/sale prices is widely studied in the real estate literature. Aside from the standard attributes (such as the number of bathrooms/bedrooms), research points to the importance of non-traditional attributes (for example, property location, being close to local amenities, schools etc.) and additional information hidden in the listing agents' descriptions which can provide further insights on sellers' motivation and house characteristics in predicting house prices (see, for example, Haag et al., 2000; Bond et al., 2002; Pryce and Oates, 2008; Nowak and Smith, 2017). In line with this literature, the installation of a solar panel can be viewed as an additional attribute or an improvement in house quality, which could increase the value of the house in the market. What is peculiar to a solar panel installation, when compared to a physical improvement such as a new bathroom or kitchen, is that it offers the possibility of energy cost savings and attracts environmentally conscious buyers (Dastrup et al., 2012). A house with a solar panel can signal environmental virtue, and, for some, this can be associated with an additional non-financial utility, increasing the value of green homes (Dastrup et al., 2012; Kahn and Kok, 2014; Sexton and Sexton, 2014).

Early evidence suggests a positive association between energy efficiency, which is broadly captured by energy bills, and residential sale prices (Halvorsen and Pollakowski, 1981; Johnson and Kaserman, 1983; Dinan and Miranowski, 1989). With the rise in eco-labelling, more recent evidence concentrates on whether these eco-labels/certificates communicating the potential energy efficiency of the houses reflect on the house values. This literature generally documents a positive relationship; for example, a more favourable energy performance according

* Corresponding author at: School of Management, Swansea University, Bay Campus, Fabian Way, Swansea SA1 8EN, United Kingdom.
E-mail addresses: ilias.asproudis@swansea.ac.uk (E. Asproudis), cigdem.gedikli@swansea.ac.uk (C. Gedikli), o.talavera@bham.ac.uk (O. Talavera), okan.yilmaz@swansea.ac.uk (O. Yilmaz).

[1] According to Renewables 2021 Global Status Report solar PV global capacity has increased from 39 gigawatts to almost 760 gigawatts from 2010 to 2020 (see: https://tinyurl.com/2untef98).

https://doi.org/10.1016/j.eneco.2024.107768
Received 17 January 2023; Received in revised form 4 July 2024; Accepted 6 July 2024
Available online 9 July 2024

to energy labels such as Energy Performance Certificate (EPC) in Europe or being certified under the LEED program or Energy Star in the US positively influences rents, vacancy rates, selling prices and external valuation (Brounen and Kok, 2011; Hyland et al., 2013; Kahn and Kok, 2014; Fuerst et al., 2015; Chegut et al., 2020; Evangelista et al., 2022). Some evidence outside the European and US contexts includes findings from Singapore by Deng et al. (2012), who note a positive selling price premium for houses with a government-designed Green Mark scheme label. By ranking houses' greenness in Beijing using a Google index, Zheng et al. (2012) point to a price premium for green buildings during the presale stage.

Therefore, there is a consensus that energy efficiency (actual or potential) and being labelled as "green" are valued favourably in the housing market. The evidence exclusively focusing on the effect of solar panels, however, remains relatively scarce and concentrates on the US and Australian real estate market (see, for example, Dastrup et al., 2012; Hoen et al., 2013; Ma et al., 2015; Qiu et al., 2017; Lan et al., 2020; Best et al., 2021). Studies rely mainly on hedonics and repeated sales approaches, and they report a positive selling price premium for houses with solar PV systems. For example, using data for California (San Diego and Sacramento) from 1997 to 2010, Dastrup et al. (2012) note a selling price premium of around 3.5% for houses with solar panels. This premium is larger in more environmentally friendly neighborhoods (captured by the number of registered Prius hybrid vehicles) and those with a larger share of college graduates. Hoen et al. (2013) provide supporting evidence that solar houses are sold at a premium ranging from \$3.9 to \$5.8 per installed watt in California. Using matching methodology on transaction and valuation data for the houses listed for sale in Arizona in 2014, more recent evidence by Qiu et al. (2017) documents that installation of solar PV is associated with an average of 15% increase in house valuation and 17% increase in the transaction price. Selling price premiums for solar houses, ranging 2.3–4.3%, are also noted for Australia (Ma et al., 2015; Lan et al., 2020).

In this paper, we contribute to the literature by stressing the value of investing in environmentally sustainable, energy-efficient properties. Our focus is on the effect of solar PV systems on sale prices in the UK housing market, for which no prior evidence exists. The installation of solar panels has been actively promoted by the UK government, especially between the years 2010 and 2019. The FiT scheme, which was introduced in April 2010, has been the most notable policy intervention to motivate individuals and businesses to opt for renewable and low-carbon electricity generation in the UK. The tariff consisted of a *generation rate,* which included payments per kWh generated by the solar system irrespective of whether it is exported or not and an *export rate,* which included payments received per kWh exported to the grid. The scheme's launch attracted many consumers, and a significant reduction in the installation cost was observed over the years.[2] Following the fall in capital costs, the UK government gradually reduced the generation tariff rates, and eventually, the scheme was closed in April 2019 for new installations.[3]

We extend the literature by utilizing novel data and a methodological approach grounded in recent advancements in machine learning, offering tools for causal inference and identifying treatment effects. While previous research has provided valuable insights, it has primarily relied on hedonic models. Despite the simplicity they offer in estimating and interpreting regression coefficients, hedonic models are susceptible to problems arising from omitted variables and are built upon strong assumptions, particularly regarding linearity parameters. This fact limits their capacity to reveal the complex non-linear relationships between property characteristics and house prices (Hong et al., 2020; Zhao and Hastie, 2021; Potrawa and Tetereva, 2022). In this paper, we apply causal machine learning tools, namely meta-learner algorithms (S-, T-, X-, R- and DR- Learners), to address these issues and advance our understanding of the causal mechanisms underlying the relationship between solar PV installation and sale prices.

We use Zoopla property data from 2012 to 2018, which is a unique data set offering information on >5 million properties advertised for sale in the UK. We then match Zoopla property data with Price Paid data from the Land Registry, which contain official records of transaction prices for each property sold in England and Wales. We capture whether a property has a solar panel installed using the textual information in the description section of the Zoopla property data.

Our results provide strong evidence that houses with solar panels bring about a sale premium that varies between 6.1% to 7.1%. To put this into context, the average selling price is £230,536 in our sample, and our findings translate into an additional £14,062 to £16,368 selling price associated with solar panels. The selling price premium we find for the UK market is larger than those documented for the US and Australia, possibly due to contextual or methodological differences. Our results remain robust when employing more standard approaches utilized in previous literature, such as hedonic pricing models or when conducting analyses on a matched sample, with the lowest estimates at 3.5%. We argue that positive returns to solar panels, a finding which is consistent across a wide range of methods, emphasize that people value green alternatives and the potential cost savings they may offer.

While this robust evidence on positive returns to solar panels is an important addition to the existing evidence base, it is crucial to acknowledge the dynamic nature of housing markets, characterized by geographical differences and influenced by changing economic conditions. As previously mentioned, significant advancements in solar technology over the years have led to notably reduced installation costs, alongside various incentives to promote solar panel adoption. While previous studies suggest that solar panels initially gained popularity in high-income and rapidly appreciating neighborhoods (Begley and Hoen, 2021; O'Shaughnessy et al., 2021), there has been a widespread increase in their adoption across different areas. Additionally, we study a period marked by significant rises in the cost of living in the UK. Thus, to delve deeper into these dynamics, we conduct further analyses examining heterogeneity based on years, regions, and price quintiles. By doing so, we aim to further contribute to the growing literature on solar panel premiums, which has thus far either overlooked these aspects or provided partial evidence.

These additional analyses demonstrate a positive but declining premium associated with solar panels, even amidst the rapidly maturing solar market characterized by significant reductions in installation costs and an increased number of houses with solar panels. Consistent with the limited evidence regarding the time trends by Begley and Hoen (2021) for California, this may suggest that solar returns now primarily reflect the cost of installation while the originality or novelty signal premiums of solar panels are declining. Our results examining regional differences and sale returns at different ends of the market show that potential cost savings, the "warm glow" associated with environmentally friendly investments, and the willingness to pay for salient property characteristics may vary in their importance in contributing to solar panel premiums. We find suggestive evidence that the premiums associated with solar panels might more likely reflect the potential cost savings they offer for poorer homebuyers. In contrast, the high-end of the sale market particularly values environmental *bona fides* and is willing to pay premiums for distinguished, salient green technologies, such as solar panels, which serve as visible indicators of the owners' commitment to environmental sustainability.

The rest of the paper is organized as follows. The next section explains the Zoopla property data and the Price Paid data by Land Registry and describes our methodology. Section 3 presents the results and

[2] According to Green Business Watch, the median cost of a 4 kW installation decreased from £20,000 to £6856 from 2010 to 2018 (see https://tinyurl.com/4k6nbjsz).

[3] The UK government introduced Smart Export Guarantee (SEG) in January 2020. Unlike the FiT scheme, SEG only pays for the electricity exported to the grid and the tariff rates change depending on the electricity supplier chosen.

interprets our findings. Section 4 concludes.

2. Data and methodology

2.1. The Zoopla and Land Registry

We match Zoopla property data[4] and Price Paid data (PPD) from Land Registry[5] to investigate the association between solar PV panels and property selling prices in the UK. Zoopla is one of the UK's leading providers of historical property listings data; it contains records of >3 million properties advertised for rent and >5 million for sale. For the purpose of our analyses, we restrict our sample to properties advertised for sale between the years 2012 and 2018.

Zoopla property data provide detailed information on properties' physical characteristics, including property type (e.g., flat; detached house, terrace house), the number of bedrooms and bathrooms, and comprehensive geographical information at the local authority level. However, the actual transaction price is not available in the data. We, therefore, match Zoopla property data with PPD from Land Registry, which contains official records of the price paid for each property sold in England and Wales. Both datasets include the postcode of the street on which the property is located. The PPD also provides a Primary Addressable Object Name (PAON), which is typically the house number. The house number for most of the listings is also available in the Zoopla property data. When the house number is missing for a listing, we extract the first numeric value from the displayable address that is available in the dataset. Since the house number and the postcode allow us to identify each property uniquely, we match Zoopla property data and PPD based on these two variables.[6] By doing so, we match 80% of the listings on Zoopla property data with the Price Paid data.[7]

We identify whether a solar panel is installed in a property using the property owners'/listing agents' written descriptions about the property available in the Zoopla property data. We categorize the properties as those with solar panels installed if their listings include the keyword "solar" or "pv panel" in the description. In determining our keywords, we examine common terms appearing before and after "solar". To ensure an accurate measure of solar panels, we also identify descriptions containing keyword combinations that might indicate other technologies, such as thermal panels or hot water heaters, or potentially indicating something other attributes with the term "solar", such as solar lights. Such descriptions are excluded when constructing our indicator of the houses with solar panels.[8] In the sample selection procedure, we exclude properties with more than three bathrooms or more than five bedrooms (each restriction corresponds to the top 1% of the distribution in the relevant variable). Also, as flats in the buildings share the same roof, we drop flats from our sample since flat owners may not be able to install solar panels. Finally, we exclude the properties whose selling prices are at the top and bottom 1% of the price distribution to remove outliers in the data.

The variable definitions and key descriptive statistics are available in Table 1. In addition to standard attributes, we account for additional quality and size-related property characteristics, such as the type of flooring, presence of a fireplace, garage, conservatory, and garden,

Table 1
Descriptive Statistics.

	Full	No solar panel	Solar panel	Difference
Log-price	12.181	12.180	12.377	−0.197***
	(0.565)	(0.566)	(0.516)	0.007
		Property characteristics		
Number of bedrooms	2.967	2.965	3.300	−0.335***
	(0.781)	(0.780)	(0.811)	0.010
Number of bathrooms	0.539	0.538	0.762	−0.224***
	(0.711)	(0.710)	(0.865)	0.009
Garden	0.831	0.830	0.909	−0.079***
	(0.375)	(0.375)	(0.288)	0.005
Garage	0.410	0.409	0.583	−0.174***
	(0.492)	(0.492)	(0.493)	0.006
Conservatory	0.108	0.108	0.215	−0.107***
	(0.311)	(0.310)	(0.411)	0.004
Flooring	0.298	0.298	0.353	−0.055***
	(0.458)	(0.457)	(0.478)	0.006
Fireplace	0.147	0.147	0.172	−0.024***
	(0.354)	(0.354)	(0.377)	0.004
		Refurbishments		
New windows	0.001	0.001	0.002	−0.001
	(0.035)	(0.035)	(0.044)	0.000
New bathroom	0.005	0.005	0.006	−0.001
	(0.071)	(0.071)	(0.078)	0.001
New kitchen	0.009	0.009	0.015	−0.006***
	(0.094)	(0.094)	(0.123)	0.001
New roof	0.002	0.002	0.004	−0.003***
	(0.040)	(0.040)	(0.067)	0.001
Refurbished (other)	0.026	0.026	0.031	−0.005**
	(0.159)	(0.159)	(0.174)	0.002
		Property age bands		
Before 1950	0.405	0.406	0.210	0.195***
	(0.491)	(0.491)	(0.407)	0.006
1950–1990	0.453	0.453	0.610	−0.157***
	(0.498)	(0.498)	(0.488)	0.006
After 1990	0.142	0.142	0.180	−0.038***
	(0.349)	(0.349)	(0.384)	0.004
		Type		
Terraced	0.346	0.347	0.166	0.181***
	(0.476)	(0.476)	(0.372)	0.006
Detached	0.532	0.532	0.631	−0.099***
	(0.499)	(0.499)	(0.483)	0.006
Bungalow	0.122	0.122	0.203	−0.081***
	(0.327)	(0.327)	(0.402)	0.004
Observations	1,347,519	1,341,289	6230	1,347,519

Note: *Solar panel* takes the value of 1 if the property listing description includes keywords identified to capture solar panels and 0 otherwise. Log-price is the natural logarithm of the sale price. The *Number of bedrooms* and *Number of bathrooms* are continuous variables. *Garden, Garage, Conservatory,* and *Fireplace* are dummy variables representing the presence of these characteristics in the property (equal to 1 if these characteristics exist and 0 otherwise). *Flooring* is a dummy variable that takes the value 1 for the presence of wooden/laminate flooring in the property and 0 otherwise. *New windows, New bathroom, New kitchen,* and *New roof* are dummy variables equal to 1 indicating whether the properties undergo these types of refurbishments and 0 otherwise. *Refurbished (other)* is a dummy capturing any other refurbishments. Property age bands are dummy variables indicating the period in which the property was built, each is a dummy variable taking the value of 1 if the property belongs to the relevant category and 0 otherwise. Property type categories are *Detached*, *Terraced,* and *Bungalow,* each is a dummy variable taking the value of 1 if the property belongs to the relevant category and 0 otherwise. Standard deviations are in parentheses. ** $p < 0.01$, *** $p < 0.001$.

[4] The data are provided by the Urban Big Data Centre (UDBC) licensed by Zoopla (2018), https://tinyurl.com/efw2p2oy.

[5] Price paid data is a publicly available data. For details see https://tinyurl.com/pf4987ke.

[6] Since properties can enter the market and be sold multiple times within the period that we are interested, we match each listing with the closest sale record on the Price Paid data.

[7] The remaining 20%, which do not match with the Price Paid data, can be due to the incorrect address entry to Zoopla. However, it is also possible that some listings can be withdrawn from the website without a sale.

[8] These additional keywords combinations include: "solar thermal panel", "thermal solar panel", "solar hot water", "hot water solar", "solar heat", "solar water heating", "solar powered water", "solar powered hot water", "solar powered heating", "solar panel heating", "solar collector", "solar vent", "solar power vent", "solar light", "solar power light", "solar pump", and "solar power pump".

along with property age bands.[9] As can be seen, on average, houses with solar panels are sold at a higher price, are bigger (have a greater number of bedrooms/bathrooms) and perform better in terms of further quality and size-related features. They are also more likely to be detached or bungalows when compared to houses without solar panels.[10]

One may argue that properties with solar panels might have a greater possibility to undergo other improvements, which may bias the selling price premium estimates (Dastrup et al., 2012). In our analyses, once again drawing on the textual information available in the description section of the Zoopla data, we identify whether the properties are renovated/refurbished, and our models control for this possibility. Accordingly, we first investigate whether the descriptions include "furbished" and "renovated". In addition, to differentiate between the extent of renovations, we expand our searches to determine specific types of renovations, such as installing new windows, bathrooms, or kitchens, by drawing on additional textual information from the descriptions.[11] Consequently, we construct a new indicator for refurbishment, distinguishing between renovations involving new windows, new bathrooms, new kitchens, and all other refurbishments. Perhaps most importantly, we also consider whether the houses underwent installation of a "new roof". This is a novel addition to the solar panel selling price premium estimations, as updating the roof might coincide with solar installation, and both can have a positive price effect. Overall, houses with solar panels are relatively more likely to undergo these various types of renovations (Table 1).

Table 2 (Panel A) shows the distribution of solar-installed houses across regions over the years under consideration. Overall, the share of houses with solar panels has increased over the years in our sample. When we look at regional differences, the highest shares of solar houses are observed in the Southwest, followed by regions located in the East Midlands and Northeastern part of the country, including Yorkshire. Excluding Greater London, the lowest shares are in Northwest, West Midlands and Wales. Finally, the lowest share of houses with solar panels is observed in the lowest price group; the share of houses with solar panels increases gradually along the selling price distribution.

2.2. Causal machine learning approach: meta-learner algorithms

It is widely noted in the real estate literature that standard hedonic pricing models may suffer from endogeneity, in particular, the omitted variables problem. One may argue that other property characteristics, for instance, whether the property is double-glazed or has alternative water heating systems, may increase energy efficiency and hence the selling prices (see, for example, Fuerst et al., 2015 for a similar discussion). Additionally, as also evident in our sample, it is possible that houses with a potentially higher market value are more likely to have solar panel systems. These possibilities persist regardless of the extent of covariates included in the hedonic pricing models. Furthermore, hedonic models tend to overlook potential non-linear relationships between property characteristics and selling prices (Zhao and Hastie, 2021). This limitation arises from the challenge of directly observing preference structures and incorporating all housing market characteristics that contribute to market complexity. In practice, various market attributes often intersect, yet conventional hedonic pricing models lack the flexibility to accommodate these multifaceted dynamics. For example, when buyer preferences exhibit non-linearities regarding a specific attribute, relying on a single regression predictor may fail to capture such nuanced factors. Similarly, including all possible interactions among property characteristics and market dynamics may lead to multicollinearity issues (Chen et al., 2020; Hong et al., 2020; Potrawa and Tetereva, 2022). In essence, certain complex aspects of selling price estimations cannot be fully addressed within the simplified assumptions of traditional hedonic pricing models.

In this paper, we therefore call for additional innovative approaches and adopt causal machine learning tools to foster an understanding of the causal effects of solar panels on sale prices. We employ *Causal ML*, a Python package developed by Uber, that offers several causal inference methods based on machine learning algorithms. For our analyses, we employ meta-learner algorithms. More specifically, we draw on S-Learner, T-Learner, X-Learner, R-Learner and DR-Learner. These techniques represent recent attempts to tailor machine learning methods for causal inquiries, providing the capacity for inference (such as controlling for confounding factors and estimating treatment effects) and extending beyond mere prediction, which has been the traditional focus of machine learning approaches.[12]

Meta-learner algorithms - meta-learners - allow estimating average treatment effects conditional on the features (i.e., covariates) based on machine learning estimators; in other words, base learners/algorithms such as linear regressions, random forests, neural networks, extreme gradient boosting (XGBoost) and so on (Künzel et al., 2019). Neither of the meta-learners is uniformly the best: their performance depends on factors such as treatment and control group size imbalances and the presence of confounding. One approach is to first use so-called base learners to estimate the conditional expectations of the outcomes separately for units under control and those under treatment and then to calculate the difference between these estimates. The S-Learner and the T-Learner that we describe below belong to this group of meta-learners. The second class of meta-learners are more sophisticated; they utilize additional information from the propensity score and are sometimes referred to as "pseudo-outcome" or "transformed outcome" methods (Jacob, 2021; Salditt et al., 2023).

We start our analysis with the most basic S-Learner algorithm that uses a single base learner and includes the treatment variable as a feature in addition to the other features. Estimating the treatment effect using S-Learner consists of two stages: assume that $Y(0)$ and $Y(1)$ are the potential outcomes (selling prices in our case) for the control and treatment groups (houses without and with solar panels), respectively.[13] In the first stage, S-Learner estimates $Y(0)$ and $Y(1)$ in a single machine learning model (hence the name "Single" or "S" Learner), which takes the treatment indicator T as a feature along with the other features X. Thus, it first estimates a combined response function (i.e., conditional mean function) as follows:

$$\mu(x,t) = [Y|X = x, T = t] \tag{1}$$

[9] The information on the size of the properties in square meters is not available in the Zoopla property data. The information on the built year is obtained from the Energy Performance of Buildings Data: England and Wales (EPC), which we merged with the Zoopla Property Data using the address (property number and postcode) information.

[10] Our categorization of property types: "Terraced houses" refers to terraced and end-terraced houses lined up in a row of uniform houses built in a continuous line. "Detached" includes houses that are detached, link-detached, and semi-detached. "Bungalows" are typically single-story detached houses, often surrounded by a garden or veranda.

[11] To identify "new windows", we conducted searches using a combination of keywords such as "new windows", "newly fitted windows", "newly fitted double glaze", and "new double glaze". Similarly, we identified houses with new bathrooms by searching for keywords like "new bathroom", "newly fitted bathroom", and "newly fitted family bathroom". Likewise, we looked for houses with new kitchens by searching for phrases such as "new kitchen", "newly fitted kitchen", "renovated kitchen", and "refurbished kitchen". Our indicators for new windows, bathrooms and kitchens are equal to 1 if the descriptions included associated keywords, and 0 otherwise.

[12] For a detailed exploration of the significance of the recent machine learning literature for economics and econometrics, see Athey and Imbens (2019).

[13] The method presented here is widely drawn from Künzel et al. (2019), Kennedy (2020), Nie & Nie and Wager (2021), and user guides for the EconML (available at https://tinyurl.com/yxfdrh7c) and Causal ML packages (available at https://tinyurl.com/dhct58w9).

Table 2
Percentage of houses with solar panels across years and regions and price groups.

Panel A: across years and regions

	2012	2013	2014	2015	2016	2017	2018	Total	Observations
East Midlands	0.30	0.41	0.55	0.72	0.72	0.86	1.24	0.63	137,742
East of England	0.21	0.33	0.38	0.44	0.68	0.84	0.63	0.47	180,357
Greater London	0.07	0.12	0.13	0.16	0.17	0.25	0.23	0.15	74,250
Northeast England	0.15	0.30	0.50	0.88	0.82	0.74	1.20	0.56	47,552
Northwest England	0.12	0.22	0.27	0.33	0.39	0.40	0.54	0.31	174,975
Southeast England	0.29	0.42	0.44	0.44	0.53	0.59	0.65	0.46	224,430
Southwest England	0.33	0.49	0.62	0.75	0.71	1.05	0.85	0.66	158,352
West Midlands	0.17	0.22	0.34	0.39	0.47	0.52	0.34	0.36	143,150
Yorkshire	0.19	0.32	0.49	0.59	0.70	1.02	0.99	0.55	141,852
Wales	0.20	0.18	0.26	0.44	0.45	0.63	0.53	0.37	64,859
Total	0.22	0.33	0.42	0.5	0.56	0.69	0.69	0.46	1,347,519
Observations	131,239	256,999	305,157	232,970	200,464	169,665	51,025	1,347,519	

Panel B: across price groups

	1st Quintile	2nd Quintile	3rd Quintile	4th Quintile	5th Quintile
	0.22	0.36	0.50	0.60	0.63
Observations	613	974	1344	1601	1698

It then estimates the treatment effect in the second stage:

$$\hat{\tau}(x) = \hat{\mu}(x, 1) - \hat{\mu}(x, 0) \tag{2}$$

where $\hat{\mu}$ is estimated by any base learner (any suitable machine learning algorithm such as supervised learning or regression estimator) using the observations both in the control and treatment groups. In our model, our treatment group consists of houses with solar panels, and the control group consists of those without solar panels. The variable capturing whether the house has a solar panel takes the value one if the listing description includes the keywords we identified to capture houses with solar panels and value zero otherwise. The features X consist of the rich set of property attributes we control for in our estimations, ranging from standard attributes such as property type and number of bedrooms/ bathrooms to additional quality and size-related characteristics, property age bands, and a detailed account of renovation indicators (Table 1). We also use location and year features, such as the county-level location (the dataset includes listings from 79 counties) and the year in which the property advertised for sale enters the market. The outcome of interest is the logarithm of the selling price of the property.

While S-Learner is a sensible starting point, it has a tendency to bias estimations toward zero. Since the S-Learner treats the treatment indicator alongside any other covariate, if the treatment is significantly weaker in comparison to the influence of other covariates in explaining the outcome, the S-Learner may overlook the treatment variable (Künzel et al., 2019).[14] Moreover, a single linear model cannot fully capture the different relevant dimensions and smoothness of features for the control and treatment groups (Alaa and Schaar, 2018). In the next set of analyses, we employ T-Learner, which attempts to overcome these problems.

As with the S-learner, the T-Learner algorithm consists of two stages, but the first stage of T-learner estimates two response functions (thus named T-) separately for the control and treatment groups:

$$\mu_0(x) = E[Y(0) | X = x] \tag{3}$$

$$\mu_1(x) = E[Y(1) | X = x] \tag{4}$$

The second stage then takes the difference between the two functions and estimates the treatment effect as follows:

$$\hat{\tau}(x) = \hat{\mu}_1(x) - \hat{\mu}_0(x) \tag{5}$$

where $\hat{\mu}_1$ and $\hat{\mu}_0$ are estimated by base learners, any potentially different suitable machine learning algorithms, using the observations in the treatment and control groups, respectively. In situations where the treatment effect is more complex, the T-learner is expected to perform well, especially when both treatment and control groups are sizable. However, if one group has limited data, the T-learner's estimates may become unstable and biased due to the potential overfitting of the small group's data, leading to differences between the functions partly influenced by random noise.

X-Learner is an expansion of T-Learner. It involves two stages and a propensity score model and is shown to be more efficient, especially when the number of units in either the treatment or control groups is notably larger than the other (see Künzel et al., 2019 for further details).[15] This is important, considering that the number of properties with solar panels is notably less than those without solar panels in our sample. X-learner estimates the average treatment effect conditional on covariates in three stages. The first stage is identical to T-Learner where response functions for the control and treatment groups are estimated separately. In the second stage, X-Learner calculates the imputed treatment effect for the treated group using the control-outcome estimator and for the control group using the treatment-outcome estimator as follows:

$$\widetilde{D}^1 = Y^1 - \hat{\mu}_0(X^1) \tag{6}$$

$$\widetilde{D}^0 = \hat{\mu}_1(X^0) - Y^0 \tag{7}$$

Using these imputed treatment effects (pseudo-outcomes) for the treated and the control groups, it then obtains two estimates of the treatment effects: one for the treatment group and one for the control group, respectively:

$$\hat{\tau}_1(x) = \mathrm{E}\left[\widetilde{D}^1 | X = x\right] \tag{8}$$

$$\hat{\tau}_0(x) = \mathrm{E}\left[\widetilde{D}^0 | X = x\right] \tag{9}$$

using any suitable machine learning algorithm. Finally, in stage 3, the

[14] However, when the Conditional Average Treatment Effect (CATE) is simple or indeed zero across various covariate value combinations, the S-learner can demonstrate effective performance (Künzel et al., 2019).

[15] It is common to observe smaller number of observations in the treatment group when compared to the control group. The small sample size may increase the chances of overfitting the treatment group when T-Learner is used. X-Learner attempts to overcome this issue by using the information from the control group in order to achieve better estimators for the treatment group and vice versa (Künzel et al., 2019).

treatment effect is estimated as a weighted average of $\hat{\tau}_0(x)$ and $\hat{\tau}_1(x)$, the estimates derived in stage 2:

$$\hat{\tau}(x) = g(x)\hat{\tau}_0(x) + (1 - g(x))\hat{\tau}_1(x) \tag{10}$$

where $g \in [0, 1]$ is the weight function utilizing propensity score for weighting. The weighting aligns the final estimate more closely with the estimated effect derived from the conditional mean function calculated for the larger group, which is expected to provide a more accurate estimate since it is trained using a greater amount of data.

R-learner, on the other hand, is a newer technique and relies on the loss function for treatment effect estimation (Nie and Wager, 2021). It represents an important attempt to mitigate the selection bias in observational studies relying mainly on Robinson's transformation (Robinson, 1988) and residualization. It is thereby an important addition to our estimations, as it helps to account for potential selection issues when assessing the impact of solar panels on house prices.[16] Briefly speaking, the R-Learner first fits the conditional mean outcomes ($\hat{m}(x)$) and propensity scores ($\hat{e}(x)$) based on cross-fitting. Then, in the second stage, it estimates the treatment effect by minimizing the so-called R-Loss function:

$$\hat{L}_n(\tau(x)) = {}^1/_n \sum_{i=1}^{n} \left(\left(Y_i - \hat{m}^{(-i)}(X_i)\right) - \left(T_i - \hat{e}^{(-i)}(X_i)\tau(X_i)\right)^2 \right. \tag{11}$$

where $\hat{m}^{(-i)}(X_i)$ and $\hat{e}^{(-i)}(X_i)$ indicate the out-of-fold predictions made without using the ith training example (see Nie and Wager (2021) p.300–301 for a more detailed mathematical derivation).

Finally, we use Doubly Robust (DR) Learner (Foster and Syrgkanis, 2019; Kennedy, 2020), which estimates the treatment effects when all the potential confounders are observed but they are high-dimensional (that is, the classical statistical tools are inapplicable due to many confounders) or their effects on the treatment and control groups cannot be estimated by parametric models effectively. As with the R-Learner, DR-Learner, developed by Kennedy (2020), falls within the recent efforts that put more emphasis on nonparametric methods and incorporate machine learning to provide more flexible conditional treatment effect estimators with robust theoretical guarantees. Accordingly, DR-Learner performs two predictions at the first stage: *i)* the outcome from the treatment and control groups and *ii)* the treatment from the controls. It then combines the predictions from these two models and runs a final stage estimation to produce a model to calculate the heterogeneous treatment effects.[17]

Overall, each meta-learner has its strengths, varying in performance based on treatment complexity, sample size balance, dimensionality, and confounding. For instance, Künzel et al. (2019) demonstrate that when sample sizes are sufficiently large, S-, T-, and X-Learners yield comparable estimates. However, in cases of imbalance between treatment and control group sizes, X-Learner shows particularly strong performance. Likewise, Nie and Wager (2021) note that all learners perform well under conditions resembling a randomized control trial, with X-, S-, and R-Learners exhibiting better mean-squared error performance. However, if there is a need to account for confounding variables before estimating treatment effects, R-Learner outperforms. Essentially, meta-learners utilizing propensity scores are preferable when confounders are a concern (Jacob, 2021; Nie and Wager, 2021; Salditt et al., 2023). Utilizing these different techniques assists in improving the accuracy of our predictions on whether a selling price premium exists for the houses with solar panels and the extent of this premium. Most importantly, they enable us to offer new evidence on the impact of having solar panels on house prices, acknowledging well-known issues such as non-linearity, selection bias, omitted variables, or unbalanced data, which traditional hedonic models used in previous research have not been able to address.

Using these tools, we estimate the treatment effect for the whole sample. Following this, we explore any regional and yearly variations and whether the selling price premium differs across the different ends of the sale market. Our base learners in T-, X-, R- and DR- Learners use XGBoost (Friedman et al., 2000; Friedman, 2001; Chen and Guestrin, 2016).[18] XGBoost is shown to be an efficient machine-learning approach and achieves the highest accuracy for all sample sizes and price groups in the rental market (see Yoshida et al., 2022).

2.3. Hedonic estimation and coarsened exact matching

We also investigate whether our results change when we employ more standard approaches. We start with a hedonic pricing model where we estimate the sale price against the property characteristics and the variable capturing the presence of solar panels in the property:

$$ln(Selling\ Price_i) = \alpha + \beta Solar_i + X_i\gamma + T_i\delta + \varepsilon_i \tag{12}$$

As with the machine learning methods, the outcome of interest is the logarithm of the selling price of the ith property. *Solar* is a dummy variable that takes the value one if the listing description includes "solar" and value zero otherwise. The vector X_i includes property characteristics described in Section 2.1. (Table 1), and T_i stands for the year and location dummies. γ and δ are the vectors capturing the shadow prices of the physical, location and year-of-listing characteristics of the houses. The coefficient β denotes the size of the association between the presence of a solar panel and the selling price.

In the next step of our analyses, we calculate the treatment effect using a matched sample where we match houses with and without solar panels using all the property attributes included in the previous models, as well as the year when the property enters the market and the county which they locate. We use coarsened exact matching (CEM). By doing so, we attempt to reduce the imbalance in the covariates between the treated and control groups. CEM uses monotonic imbalance bounding (Iacus et al., 2012) and is shown to be more effective in achieving lower levels of imbalance, model dependence and bias than other commonly used matching techniques, such as propensity score matching (Blackwell et al., 2010; Iacus et al., 2012; King and Nielsen, 2019). Once the data are matched, we then calculate the average treatment of solar panels by including CEM weights in linear regression.

3. Results

3.1. Results from the meta-learner algorithms

The results from the various meta-learner algorithms are presented in Table 3. The average treatment effect calculated using S-Learner algorithm is 0.055, indicating that houses with solar panels command 5.5% selling price premium. When we use more sophisticated meta-learners, the premiums associated with solar panels increase to >6%, ranging between 6.1 (X-Learner) percent to 7.1% (DR-Learner).

[16] The concept of achieving Neyman orthogonality through a residuals-on-residuals method has its roots in econometrics. The R-Learner draws on Robinson (1998) transformation and the broader literature on semiparametric efficiency and orthogonal moments, thereby facilitating flexible treatment effect estimation through modern machine learning techniques and emphasizing the model's focus on "residualization" (Nie & Wagner, 2021). To account for selection bias, R-Learner regresses the residuals obtained from the regression of the outcome variable Y on covariates X against the residuals derived from regressing the treatment variable D on X, and weighted by the squared residuals (Jacob, 2021).

[17] Kennedy (2020) notes that DR-Learner can identify treatment effects even when the causal assumptions regarding "no unmeasured confounding" are not met.

[18] Gradient boosting provides a prediction model based on an ensemble of weak prediction models such as decision trees.

Table 3
Treatment effects based on meta-learner estimators.

Meta learners	Treatment effect	%95 CI
S-learner	0.055	(0.047), (0.063)
T-learner	0.068	(0.062), (0.074)
X-learner	0.061	(0.055), (0.067)
R-learner	0.063	(−0.016), (0.117)
DR-learner	0.071	(0.065), (0.078)

Note: The outcome of interest is the natural logarithm of the sale price. The treatment group includes properties with solar panels; the treatment indicator takes the value of 1 if the property description includes keywords identified to capture solar panels and 0 otherwise. The models comprise of all the features (covariates) presented in Table 1. Additionally, they include the county where the property locates (79 counties in total) and the year the property enters the market. The R-Learner standard errors are bootstrapped.

These results are consistent with the literature documenting a positive selling price premium for houses with solar panels. However, the premium we find is relatively larger than the ones noted for California by Dastrup et al. (2012) and Hoen et al. (2013) and for Australia by Ma et al. (2015) and Lan et al. (2020). Several factors might have contributed to this difference. First, it may simply be the case that we focus on the UK market, and there may be higher returns on solar panels in the UK. However, the difference may also be related to the fact that we employ causal machine-learning tools which have not been used in the relevant literature before. Meta-learners, being better equipped to account for the diverse dimensions and smoothness of features for both control and treatment groups compared to single linear models, as well as their ability to capture non-linearity in the relationship between property characteristics and house prices, are more likely to produce more accurate predictions. This reduces the risk of underestimating the true effect of solar panels by increasing the explained variability in house prices attributable to them.[19]

Our further analyses, which explore differences across years, regions, and price quintiles, offer valuable additional insights into the positive selling price premiums associated with solar panels. Firstly, our results reveal a declining trend in returns to solar panels over the years. However, despite this decline, a positive premium at around 2–4% persists, except for a slightly negative return observed in 2018 based on the DR-Learner estimates (see Table 4, Panel A). This declining trend in returns to solar panels is expected, given the significant decrease in installation costs over the years and the growing number of houses with solar panels.

The existing evidence on solar panel selling price premiums over time is limited. To the best of our knowledge, only one study, Begley and Hoen (2021), has investigated the changes in premiums associated with solar panels over time in California, covering the years between 2011 and early 2020. Similar to our findings, this study documents a declining trend, which is attributed to the decrease in replacement costs and the overall maturation of the solar market. Expanding on related research indicating that early solar panel adopters are motivated by environmental concerns and novelty, while later adopters prioritize economic gains and are often less wealthy (O'Shaughnessy et al., 2021; Palm, 2020), the authors note that as solar panels become more widespread, the premiums associated with early adoption, originality, and novelty diminish. Instead, they increasingly reflect the cost of installation and the potential energy savings they provide. Our findings suggest that similar dynamics may be at play in the UK housing market.

Next, we investigate whether there are any regional variations in returns to solar panels (see Table 4, Panel B). The highest returns are observed in Wales and Yorkshire and the Humber, followed by comparable estimates for the Northwest and Southwest regions. While we cannot directly test the reasons behind the regional differences, several potential mechanisms may be speculated, operating independently or simultaneously. Firstly, considering the positive association between energy efficiency and house prices documented by the previous literature, one might expect the premiums associated with solar panels to be higher in regions with greater solar radiance, assuming all other factors remain constant. While this prior could account for the relatively higher premiums in the Southwest, it does not explain the highest returns we observe for the solar houses in Wales and Yorkshire and the Humber, or in the Northwest, despite the poorer solar radiance in these regions.[20] Drawing on the saliency and novelty signal premiums of solar panels (Begley and Hoen, 2021; Shen et al., 2021), another possibility could be that in regions where houses with solar panels are more common, the novelty and uniqueness associated with solar panels may be less pronounced, leading to lower premiums. Conversely, in regions with fewer solar installations, the novelty factor might increase premiums. However, even though we see high premiums in Wales and the Northwest where the incidence of solar houses is lower, our data do not fully support this argument either, considering the high returns in Yorkshire and the Humber and the Southwest despite their high rates of solar houses. Instead, the highest solar returns in Wales, Yorkshire and the Humber, and the Northwest may be due to the potential cost savings associated with solar panels. Given the relatively higher poverty rates, consumers in these regions may particularly value the chance to save on energy expenses.[21]

Although these mechanisms may operate simultaneously, the premiums we identify might underscore the importance low-income buyers attach to potential energy savings due to their higher energy burden (i.e., energy costs relative to income). Conversely, the saliency argument may apply to high-income buyers, as noted in previous research (see Shen et al., 2021). In this sense, our analyses focusing on solar panel returns at different price quintiles can be informative.

The effect of solar panels on house prices is more pronounced in the lowest and highest price quintiles, with the largest premiums observed in the latter (Table 4, Panel C). The premiums we observe for the lowest price quintile support our argument regarding potential energy savings and the significance of cheaper running costs for buyers with lower incomes, given their higher energy burden. Additionally, it is plausible that any price increase associated with the installation of a solar panel is likely to have a greater impact on houses in the lowest price band, making solar panel installation a valuable investment. For the highest quintile, the role of cheaper running costs may be less influential in driving the price premium. The premiums observed in the high-end of the market may instead reflect the "warm glow" associated with investing in environmentally friendly properties, as well as the "green" signaling and conspicuous conservation effect (Dastrup et al., 2012; Sexton and Sexton, 2014).[22] Wealthier buyers may place a higher value on environmental bona fides and be willing to pay more for properties that showcase them. They may pay premiums for distinguished and salient green technologies, such as solar panels, which serve as visible indicators of the owners' commitment to environmental sustainability (Dastrup et al., 2012; Sexton and Sexton, 2014; Shen et al., 2021).

[19] For the UK residential market, there is prior evidence for the effect of EPC ratings on the selling prices; Fuerst et al. (2015) find that properties in EPC Bands A and B ask for 5% selling price premium while for those in Bands C, the premium is 1.8%. Although direct comparison is still not achievable, our findings signal a slightly more pronounced return to green properties.

[20] Data on solar radiance in the UK can be found at https://tinyurl.com/ymhc2t3k.

[21] Data on UK poverty rates by region can be found at https://tinyurl.com/27e5u65w.

[22] The conspicuous conservation theory (Sexton and Sexton, 2014) illustrates that, amidst rising concerns over environmental damage and global climate change, individuals seek status through displaying austerity. Consequently, consumers may make costly investments to demonstrate prosocial behavior concerning environmental protection, a concept referred to as conspicuous conservation.

Table 4
Treatment effects based on meta-learner estimators by years, regions, and price groups.

	S-learner	T-learner:	X-learner	R-learner	DR-learner
Panel A: Years					
2012	0.099 [0.060, 0.138]	0.073 [0.058, 0.087]	0.083 [0.068, 0.097]	0.063 [−0.008, 0.170]	0.072 [0.041, 0.103]
2013	0.085 [0.063, 0.108]	0.108 [0.097, 0.120]	0.091 [0.080, 0.103]	0.092 [0.007, 0.163]	0.114 [0.097, 0.132]
2014	0.062 [0.044, 0.080]	0.064 [0.053, 0.074]	0.055 [0.045, 0.065]	0.057 [−0.019, 0.112]	0.058 [0.044, 0.072]
2015	0.060 [0.042, 0.077]	0.066 [0.056, 0.075]	0.053 [0.044, 0.063]	0.048 [−0.016, 0.106]	0.060 [0.046, 0.074]
2016	0.056 [0.039,0.074]	0.077 [0.067, 0.087]	0.054 [0.044, 0.065]	0.057 [0.002, 0.121]	0.062 [0.048, 0.076]
2017	0.030 [0.012, 0.048]	0.034 [0.024, 0.043]	0.022 [0.012, 0.031]	0.025 [−0.030, 0.067]	0.040 [0.026, 0.054]
2018	0.022 [−0.008, 0.052]	0.008 [−0.002, 0.019]	0.027 [0.017, 0.037]	0.038 [−0.026, 0.074]	−0.011 [−0.038, 0.017]
Panel B: Regions					
East Midlands	0.055 [0.036, 0.073]	0.062 [0.054, 0.069]	0.056 [0.049, 0.064]	0.053 [0.005, 0.102]	0.034 [0.021, 0.048]
East of England	0.055 [0.035, 0.076]	0.066 [0.058, 0.073]	0.048 [0.041, 0.055]	0.059 [−0.010, 0.110]	0.064 [0.049, 0.079]
Northeast England	0.049 [0.010, 0.087]	0.045 [0.036, 0.054]	0.028 [0.019, 0.037]	0.033 [−0.019, 0.080]	0.072 [0.041, 0.104]
Northwest England	0.065 [0.036, 0.094]	0.053 [0.045, 0.061]	0.075 [0.067, 0.083]	0.064 [−0.029, 0.149]	0.044 [0.023, 0.066]
Southeast England	0.043 [0.026, 0.060]	0.046 [0.040, 0.053]	0.041 [0.034, 0.047]	0.043 [−0.009, 0.087]	0.050 [0.037, 0.062]
Southwest England	0.075 [0.058, 0.093]	0.067 [0.060, 0.073]	0.065 [0.058, 0.072]	0.064 [0.020, 0.107]	0.072 [0.059, 0.084]
West Midlands	0.032 [0.006, 0.057]	0.046 [0.038, 0.054]	0.038 [0.030, 0.046]	0.037 [−0.033, 0.093]	0.037 [0.018, 0.056]
Yorkshire	0.058 [0.030, 0.086]	0.109 [0.098, 0.120]	0.091 [0.080, 0.103]	0.088 [0.019, 0.188]	0.137 [0.116, 0.157]
Greater London	0.070 [0.019, 0.121]	0.074 [0.066, 0.082]	0.046 [0.039, 0.054]	0.049 [−0.043, 0.152]	0.097 [0.055, 0.139]
Wales	0.101 [0.060, 0.141]	0.115 [0.106, 0.123]	0.077 [0.068, 0.085]	0.049 [0.003, 0.120]	0.121 [0.090, 0.152]
Panel C: Price Groups by Quintiles					
Quintile 1	0.006 [−0.009, 0.021]	0.010 [0.003, 0.017]	0.013 [0.006, 0.020]	0.011 [−0.054, 0.066]	0.012 [0.000, 0.024]
Quintile 2	0.005 [0.000, 0.011]	0.007 [0.004, 0.010]	0.007 [0.004, 0.010]	0.008 [−0.013, 0.026]	0.000 [−0.004, 0.005]
Quintile 3	0.004 [−0.001, 0.009]	0.000 [−0.003, 0.002]	−0.001 [−0.004, 0.002]	0.000 [−0.013, 0.018]	−0.001 [−0.005, 0.003]
Quintile 4	0.005 [0.000, 0.009]	0.003 [0.000, 0.006]	0.004 [0.002, 0.007]	0.003 [−0.011, 0.021]	0.000 [−0.003, 0.004]
Quintile 5	0.028 [0.017, 0.038]	0.030 [0.024, 0.036]	0.027 [0.021, 0.033]	0.028 [−0.018, 0.071]	0.019 [0.011, 0.027]

The outcome of interest is the natural logarithm of the sale price. The treatment group includes properties with solar panels; the treatment indicator takes the value of 1 if the property description includes keywords identified to capture solar panels and 0 otherwise. All the models presented in Panels A, B and C include the features (covariates) presented in Table 1 along with the counties where the properties are located (79 counties in total). Models in Panels B and C include the year the property enters the market. The R-Learner standard errors are bootstrapped.

3.2. Results from the hedonic pricing model and the matched sample

The results from the hedonic pricing model estimated by OLS are presented in Table 5. In the first specification (Column 1), we only include the variable indicating the presence of solar panels while controlling for the year the properties enter the market and the location (i.e., the counties). The coefficient for "Solar" in this specification is 0.210, suggesting that houses with solar panels are associated with 21% higher sale prices in comparison to those without solar panels. When the model controls for quality and size-related property characteristics and refurbishments,), this premium goes down to 7.5% (Column 2). Once all the property characteristics are controlled for, including the type of the property and the property age bands, the size of the solar premium drops to 5.6% (Column 3). The selling price premium in Column 3 is close to the one estimated using S-Learner, as anticipated. This similarity is affirming, as S-Learner employs a single linear regression model where the treatment indicator is included as a feature along with other covariates.

In the next step of our analyses, we match the houses, which are almost identical across all the covariates that we can control for aside from the presence of a solar panel. We then compare the selling price of pairs of properties with and without solar panels. The average treatment calculated using the data matched by CEM are shown in Table 6. Results from the matched sample document a positive yet relatively smaller selling price premium for the houses with solar panels. The average treatment effect for houses with solar panels is 0.035, indicating a 3.5% sale premium.

Therefore, the results are somehow sensitive to the methods we use. Overall, it appears that the results from the meta-learners are somewhat larger compared to the hedonic estimates based on the unmatched sample (Table 5, Column 3), as well as the sample matched using CEM (Table 6). As indicated previously, compared to hedonic models, meta-

Table 5
Hedonic price regression results.

	(1)	(2)	(3)
	Log-price	Log-price	Log-price
Solar panel	0.210***	0.075***	0.056***
	(0.005)	(0.004)	(0.004)
Property characteristics			
Number of bedrooms		0.274***	0.259***
		(0.000)	(0.000)
Number of bathrooms		0.030***	0.023***
		(0.000)	(0.000)
Garden		0.007***	0.003***
		(0.001)	(0.001)
Garage		0.162***	0.093***
		(0.001)	(0.001)
Conservatory		0.063***	0.029***
		(0.001)	(0.001)
Flooring		−0.002***	0.007***
		(0.001)	(0.001)
Fireplace		0.066***	0.066***
		(0.001)	(0.001)
Refurbishments			
New windows		0.046***	0.045***
		(0.008)	(0.008)
New bathroom		0.011**	0.018***
		(0.004)	(0.004)
New kitchen		−0.005	0.003
		(0.003)	(0.003)
New roof		0.015	0.023**
		(0.008)	(0.007)
Refurbished (other)		0.094***	0.099***
		(0.002)	(0.002)
Property age bands			
1950–1990			−0.045***
			(0.001)
After 1990			0.059***
			(0.001)
Type			
Detached			0.209***
			(0.001)
Bungalow			0.359***
			(0.001)
Observations	1,347,519	1,347,519	1,347,519
Adjusted R^2	0.450	0.646	0.686

Note: The dependent variable *Log-price* is the natural logarithm of the sale price. *Solar panel* takes the value of 1 if the property listing description includes keywords identified to capture solar panels and 0 otherwise. The models comprise of all the features (covariates) presented in Table 1. All the specifications include the county where the property is located (79 counties in total) and the year the property enters the market. Robust standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 6
Average Treatment Effect (ATT) based on the sample matched using Coarsened Exact Matching (CEM).

	CEM
ATT	0.035***
	(0.008)
Treated (solar = 1)	6230
Untreated (solar = 0)	1,341,289
Matched treated	5502
Unmatched treated	728

Note: Properties with and without solar panels are matched using all the covariates described in Table 1, in addition to the 79 counties and year dummies. Standard errors are in parenthesis. ***Statistically significant at 1%.

learner algorithms are better able to account for issues related to treatment complexity, high dimensionality, the non-linear relationship between property characteristics and house prices and the presence of confounding.

Controlling for all relevant factors poses a significant challenge, and hedonic estimates can be subject to both upward and downward biases depending on the relationship between omitted variables and house prices, as well as their relationship to the presence of solar panels. While we control for most of the quality, size and renovation-related factors, thereby reducing the possibility of upward biased hedonic estimates, there may still be unobservable factors, such as buyer's or seller's characteristics, or neighborhood and local market characteristics, which we cannot fully account for, even with detailed locational controls. If these remaining unobservable elements are positively correlated with house prices and positively related to solar panels, hedonic estimates may be upward biased. On the contrary, if the unobservables are negatively correlated with price but positively correlated with having solar panels (or vice versa), then hedonic results could suffer from downward bias. Most importantly, if there indeed exists a non-linear relationship between house prices and property characteristics, as noted by prior research, the inability of hedonic models to fully capture this may result in underestimation of the effect, i.e., downward biased estimates of the true relationship between solar panels and house prices. The larger estimates produced by meta-learners can be understood within the context of these possibilities. By accounting for non-linearity and moving beyond linear models, which cannot fully capture the relevant dimensions and smoothness of features for both control and treatment groups, meta-learners can enhance the explained variability in house prices attributable to solar panels.

Nonetheless, it is important to note that the positive price premium associated with solar panels remains consistent across all estimations, highlighting the robust effect of solar panels on house prices. For comparison purposes, we also looked at the yearly and regional differences and the changing returns across the price quantiles. Overall, results from these attempts were qualitatively similar to the ones we received from meta-learner algorithms (see Appendix Table A1).

4. Conclusion

In this paper, we explore the returns to solar panels by focusing on their effect on the sale prices in the UK housing market. Drawing on the most recent and encouraging developments in causal machine learning methods, we attempt to estimate the effect of solar panels on selling prices. Utilizing these additional tools alongside the traditional methods is worthwhile and can provide further insights into the value of investing in energy-efficient, green properties.

Using Zoopla property data with a rich set of information for almost 1.5 million houses advertised for sale in the UK, we distinguish between houses with and without solar panels. We rely on meta-learner algorithms, which are a class of causal machine-learning techniques. Accordingly, we estimate the average treatment effect of solar panels using S-Learner, T-Learner, X-Learner, R-Learner and DR-Learner. Apart from the S-Learner estimation of around 5.5%, which is similar to a hedonic model with full controls, our results from T-Learner, X-Learner, R-Learner and DR-Learner meta-learners show a selling price premium above 6% (ranging from 6.1% to 7.1%) associated with houses with solar panels. Considering that the average selling price is £230,536 in our sample, the selling price premium of around 6.1% to 7.1% corresponds to an additional £14,062 to £16,368 selling price for the houses with solar panels. Our results are robust when we repeat the analysis using the hedonic pricing model or when we estimate the average treatment effect on a matched sample using CEM, although the latter points to a relatively smaller selling premium.

By expanding the previous literature with a novel methodological approach, our results further stress that energy efficiency and/or environmentally conscious choices are capitalized into home values. We further contribute to the existing literature by investigating whether the selling price premiums associated with solar panels are sustained over time. Our findings indicate a declining premium associated with solar panels over the years, consistent with reduced installation costs and the increasing share of solar panels in a rapidly maturing market. However, despite this decline implying lower returns for sellers, a positive premium associated with solar panels persists over the years in the UK housing market. This evidence underscores the continued importance of solar panels as an investment, with an increasing number of adopters. From a policy perspective, this robust evidence can encourage consumers and potential sellers to choose sustainable energy sources that are good for the environment.

Our results also provide additional insights into the varying effects of solar panels on selling prices across different segments of the housing market and region, which could be of interest to policymakers. We demonstrate that the premiums associated with solar panels tend to be highest in regions with relatively higher poverty rates and are more pronounced in the lowest and highest price quintiles. While the higher returns at the high-end of the market might reflect status-seeking behavior through "green" signaling and the willingness of high-income buyers to pay for additional salient property characteristics, the premiums observed in the lowest price quintiles and poorer regions are more likely driven by the value that lower-income house buyers place on potential cost savings in energy expenses. This should motivate governments to provide further incentives to reduce the cost of going green while reaching targets toward tackling climate change and carbon externalities from energy consumption. Although the cost of solar PV installation has decreased noteworthily in the UK over the years, continuation of policies aimed at reducing the cost of solar panel installation remain worthwhile, especially in neighborhoods where the shares of solar panels are lower and poverty rates are higher. Equitable energy policies, prioritizing expanding access to energy-efficient alternatives like solar panels, are vital for low-income households, given their higher energy burden.

CRediT authorship contribution statement

Elias Asproudis: Writing – review & editing, Writing – original draft, Software, Resources, Formal analysis, Conceptualization. **Cigdem Gedikli:** Writing – review & editing, Writing – original draft, Software, Resources, Formal analysis, Conceptualization. **Oleksandr Talavera:** Writing – review & editing, Writing – original draft, Software, Resources, Formal analysis, Conceptualization. **Okan Yilmaz:** Writing – review & editing, Writing – original draft, Software, Resources, Formal analysis, Conceptualization.

Declaration of competing interest

None.

Appendix A. Appendix

Table A1
Marginal effects calculated for having solar panels installed interacted with year, region and price quintile dummies.

Panel A: Interaction with year dummies	Dependent variable: Log(sale price)
2012	0.095***
	(0.020)
2013	0.084***
	(0.012)
2014	0.064***
	(0.009)
2015	0.057***
	(0.009)
2016	0.047***
	(0.009)
2017	0.034***
	(0.009)
2018	0.023
	(0.015)

Panel B: Interaction with region dummies	
East Midlands	0.063***
	(0.010)
East of England	−0.008
	(0.013)
Greater London	0.029
	(0.028)
Northeast England	0.052*
	(0.023)
Northwest England	0.094***
	(0.016)
Southeast England	0.004
	(0.010)
Southwest England	0.025**
	(0.009)
West Midlands	0.048***
	(0.014)
Yorkshire	0.071***

(*continued on next page*)

Table A1 (*continued*)

Panel B: Interaction with region dummies	
	(0.014)
Wales	0.118***
	(0.023)
Panel C: Interaction with price quintiles dummies	
1st quintile	0.019*
	(0.008)
2nd quintile	−0.003
	(0.003)
3rd quintile	−0.000
	(0.002)
4th quintile	0.006*
	(0.003)
5th quintile	0.020***
	(0.005)
Observations	1,347,519

Note: The dependent variable *Log-price* is the natural logarithm of the sale price. *Solar panel* takes the value of 1 if the property listing description includes keywords identified to capture solar panels and 0 otherwise. The models comprise of all the features (covariates) presented in Table 1. All the specifications include the county where the property is located (79 counties in total) and the year the property enters the market. Robust standard errors are in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.eneco.2024.107768.

References

Alaa, A., Schaar, M., 2018. Limits of estimating heterogeneous treatment effects: guidelines for practical algorithm design. In: International Conference on Machine Learning, pp. 129–138.
Athey, S., Imbens, G.W., 2019. Machine learning methods that economists should know about. Ann. Rev. Econom. 11, 685–725.
Begley, J., Hoen, B., 2021. Solar at High Noon: Solar Home Premiums in a Rapidly Maturing Market. Available at SSRN. https://ssrn.com/abstract=3894549.
Best, R., Burke, P.J., Nepal, R., Reynolds, Z., 2021. Effects of rooftop solar on housing prices in Australia. Aust. J. Agric. Resour. Econ. 65, 493–511.
Blackwell, M., Iacus, S., King, G., Porro, G., 2010. Cem: coarsened exact matching in Stata. Stata J. 9 (4), 524–546.
Bond, M.T., Seiler, V.L., Seiler, M.J., 2002. Residential real estate prices: a room with a view. J. Real Estate Res. 23 (1–2), 129–138.
Brounen, D., Kok, N., 2011. On the economics of energy efficiency in the housing market. J. Environ. Econ. Manag. 62 (2), 166–179.
Chegut, A., Eichholtz, P., Holtermans, R., et al., 2020. Energy efficiency information and valuation practices in rental housing. J. Real Estate Financ. Econ. 60, 181–204.
Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. arXiv: 1603.02754. https://doi.org/10.48550/arXiv.1603.02754.
Chen, L., Yao, X., Liu, Y., et al., 2020. Measuring impacts of urban environmental elements on housing prices based on multisource data - A case study of Shanghai, China. ISPRS Int. J. Geo Inf. 9 (2), 106.
Dastrup, S.R., Zivin, J.G., Costa, D.L., Khan, M.E., 2012. Understanding the solar home price premium: electricity generation and "green" social status. Eur. Econ. Rev. 56, 961–973.
Deng, Y., Zhiliang, L., Quigley, J.M., 2012. Economic returns to energy efficient investments in the housing market: evidence from Singapore. Reg. Sci. Urban Econ. 42 (3), 506–515.
Dinan, T., Miranowski, J., 1989. Estimating the implicit price of energy efficiency improvements in the residential housing market: A hedonic approach. J. Urban Econ. 25 (1), 52–67.
Evangelista, R., e Silva, J.A., Ramalho, E.A., 2022. How heterogeneous is the impact of energy efficiency on dwelling prices? Evidence from the application of the unconditional quantile hedonic model to the Portuguese residential market. Energy Econ. 109, 105955.
Foster, D., Syrgkanis, V., 2019. Orthogonal statistical learning arXiv:1901.09036. http://arxiv.org/abs/1901.09036.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232.
Friedman, J., Hastie, T., Tibshirani, R., 2000. Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). Ann. Stat. 28 (2), 337–407.
Fuerst, F., McAllister, P., Nanda, A., Wyatt, P., 2015. Does energy efficiency matter to homebuyers? An investigation of EPC ratings and transaction prices in England. Energy Econ. 48, 145–156.
Haag, J.T., Rutherford, R.C., Thomson, T.A., 2000. Real estate agent remarks: help or hype? J. Real Estate Res. 20 (1), 205–215.
Halvorsen, R., Pollakowski, H., 1981. The effects of fuel prices on house prices. Urban Stud. 18, 205–211.
Hoen, B., Wiser, R., Thayer, M., Cappers, P., 2013. Residential photovoltaic energy systems in California: the effect on home sales prices. Contemp. Econ. Policy 31 (4), 708–718.
Hong, J., Choi, H., Kim, W.- sung., 2020. A house price valuation based on the random forest approach: the mass appraisal of residential property in South Korea. Int. J. Strateg. Prop. Manag. 24 (3), 140–152.
Hyland, M., Lyons, R., Lyons, S., 2013. The value of domestic building energy efficiency: evidence from Ireland. Energy Econ. 40, 943–952.
Iacus, S., King, G., Porro, G., 2012. Causal inference without balance checking: coarsened exact matching. Polit. Anal. 20 (1), 1–24.
Jacob, D., 2021. CATE meets ML. Digit. Finance 3, 99–148.
Johnson, R., Kaserman, D., 1983. Housing market capitalization of energy-saving durable good investments. Econ. Inq. 21 (3), 374–386.
Kahn, M.E., Kok, N., 2014. The capitalization of green labels in the California housing market. Reg. Sci. Urban Econ. 47, 25–34.
Kennedy, E.H., 2020. Optimal doubly robust estimation of heterogeneous causal effects arXiv:2004.14497. https://arxiv.org/abs/2004.14497.
King, G., Nielsen, R., 2019. Why propensity scores should not be used for matching. Polit. Anal. 27 (4), 435–454.
Künzel, S.R., Sekhon, J.S., Bickel, P.J., Yu, B., 2019. Metalearners for estimating heterogeneous treatment effects using machine learning. Proc. Natl. Acad. Sci. 116 (10), 4156–4165.
Lan, H., Gou, Z., Yang, L., 2020. House price premium associated with residential solar photovoltaics and the effect from feed-in tariffs: a case study of Southport in Queensland, Australia. Renew. Energy 161, 907–916.
Ma, C., Polyakov, M., Pandit, R., 2015. Capitalization of residential solar photovoltaic systems in Western Australia. Aust. J. Agric. Resour. Econ. 60, 366–385.
Nie, X., Wager, S., 2021. Quasi-oracle estimation of heterogeneous treatment effects. Biometrika 108 (2), 299–319.
Nowak, A., Smith, P., 2017. Textual analysis in real estate. J. Appl. Econ. 32 (4), 896–918.
O'Shaughnessy, E., Barbose, G., Wiser, R., et al., 2021. The impact of policies and business models on income equity in rooftop solar adoption. Nature Energy Nature 6 (1), 84–91.
Palm, A., 2020. Early adopters and their motives: differences between earlier and later adopters of residential solar photovoltaics. Renew. Sust. Energ. Rev. 133, 110142.
Potrawa, T., Tetereva, A., 2022. How much is the view from the window worth? Machine learning-driven hedonic pricing model of the real estate market. J. Bus. Res. 144, 50–65.
Pryce, G., Oates, S., 2008. Rhetoric in the language of real estate marketing. Hous. Stud. 23 (2), 319–348.
Qiu, Y., Wang, Y.D., Wang, J., 2017. Soak up the sun: impact of solar energy systems on residential home values in Arizona. Energy Econ. 66, 328–336.
Robinson, P.M., 1988. Root-N-consistent semiparametric regression. Econometrica 56, 931–954.

Salditt, M., Eckes, T., Nestler, S.A., 2023. Tutorial introduction to heterogeneous treatment effect estimation with Meta-learners. Adm. Policy Ment. Health Ment. Health Serv. Res. https://doi.org/10.1007/s10488-023-01303-9.

Sexton, S.E., Sexton, A.L., 2014. Conspicuous conservation: the Prius halo and willingness to pay for environmental bona fides. J. Environ. Econ. Manag. 67 (3), 303–317.

Shen, X., Liu, P., Qiu, Y., et al., 2021. Estimation of change in house sales prices in the United States after heat pump adoption. Nat. Energy 6, 30–37.

Yoshida, T., Murakami, D., Seya, H., 2022. Spatial prediction of apartment rent using regression-based and machine learning-based approaches with a large dataset. J. Real Estate Financ. Econ. https://doi.org/10.1007/s11146-022-09929-6.

Zhao, Q., Hastie, T., 2021. Causal interpretations of black-box models. J. Bus. Econ. Stat. 39 (1), 272–281.

Zheng, S., Wu, J., Kahn, M.E., Deng, Y., 2012. The nascent market for green real estate in Beijing. Eur. Econ. Rev. 56 (5), 974–984.

Zoopla, 2018. Zoopla Limited. Economic and Social Research Council. Zoopla Property Data [Data Collection]. University of Glasgow - Urban Big Data Centre.